MODES AND TEMPOS IN THE EVOLUTION OF NUCLEAR RIBOSOMAL DNA: NEW CHARACTERS FOR EVOLUTIONARY STUDIES AND NEW MARKERS FOR GENETIC AND POPULATION STUDIES[1]

Richard A. Jorgensen[2,3] *and*
Paul D. Cluster[2]

ABSTRACT

The tempo of evolutionary change determines in what manner any class of characters is informative in evolutionary studies and at which taxonomic levels. Here we describe and summarize some fundamental features of the evolution of the DNA sequences that encode ribosomal RNA genes in the nuclear genome of higher plants. By analyzing a sample of angiosperm species having known phylogenetic relationships at five different taxonomic levels ranging from the intraspecific to the interfamilial, we show that plant ribosomal DNA determines at least eleven classes of characters that can be distinguished by comparisons at the DNA level. These classes are temporal and physical subsets of three basic modes of variation: length variation, single base pair substitution, and DNA modification. We also discuss the impact of length variants on population genetic studies and the implications of these studies for understanding the molecular mechanisms of rDNA evolution.

Because DNA is the richest and most unambiguous source of genetic variability, information on its evolution is fundamentally important to evolutionary biology. Research into the evolution of DNA is still in its infancy, and workers studying DNA variation are still faced with (1) cataloging both the classes of DNA sequences (characters) that are found in the genomes of various organisms and the ways (modes) in which these characters vary among organisms, and (2) measuring the approximate rate (tempo) of change in the different character classes. From this fundamental information it is possible to begin to ascertain at which phylogenetic level a particular character is useful in reconstructing phylogeny. Here we examine modes and tempos of evolution in nuclear-encoded ribosomal DNA. The available technology allows us to estimate the tempo of evolution in three basic modes: length variation, base pair substitution, and nucleotide modification.

PHYSICAL AND GENETIC DESCRIPTION OF rDNA

Ribosomal DNA, or rDNA, is the set of DNA sequences that directs the synthesis of ribosomal RNA. Each haploid nuclear genome of a higher plant cell typically contains 1,000 to 10,000 copies of ribosomal DNA (Ingle et al., 1975), a range roughly twenty-fold higher than in animal genomes. Copies of rDNA exist in long tandem arrays at one or a few chromosomal locations. Within a species, the number of copies of rDNA varies by as much as four-fold (Cullis & Davies, 1975; Long & Dawid, 1980). Unequal crossing-over is one likely mech-

[1] *Editor's note: This and the following article by H. J. Price are based upon the authors' presentations at the 1982 A.I.B.S. meetings as part of a symposium originally planned for publication in the* Annals of the Missouri Botanical Garden. *After the manuscripts were prepared, plans were changed, and the symposium was never published. Publication here of the 1987 Missouri Botanical Garden Annual Systematics Symposium, dealing with the same subject matter, provides an opportunity to present these updated papers as complements to the more recent symposium.*

[2] *Department of Genetics, University of California, Davis, California 95616, U.S.A.*

[3] *Present address: Advanced Genetic Sciences, Inc., 6701 San Pablo Avenue, Oakland, California 94608, U.S.A.*

ANN. MISSOURI BOT. GARD. 75: 1238–1247. 1988.

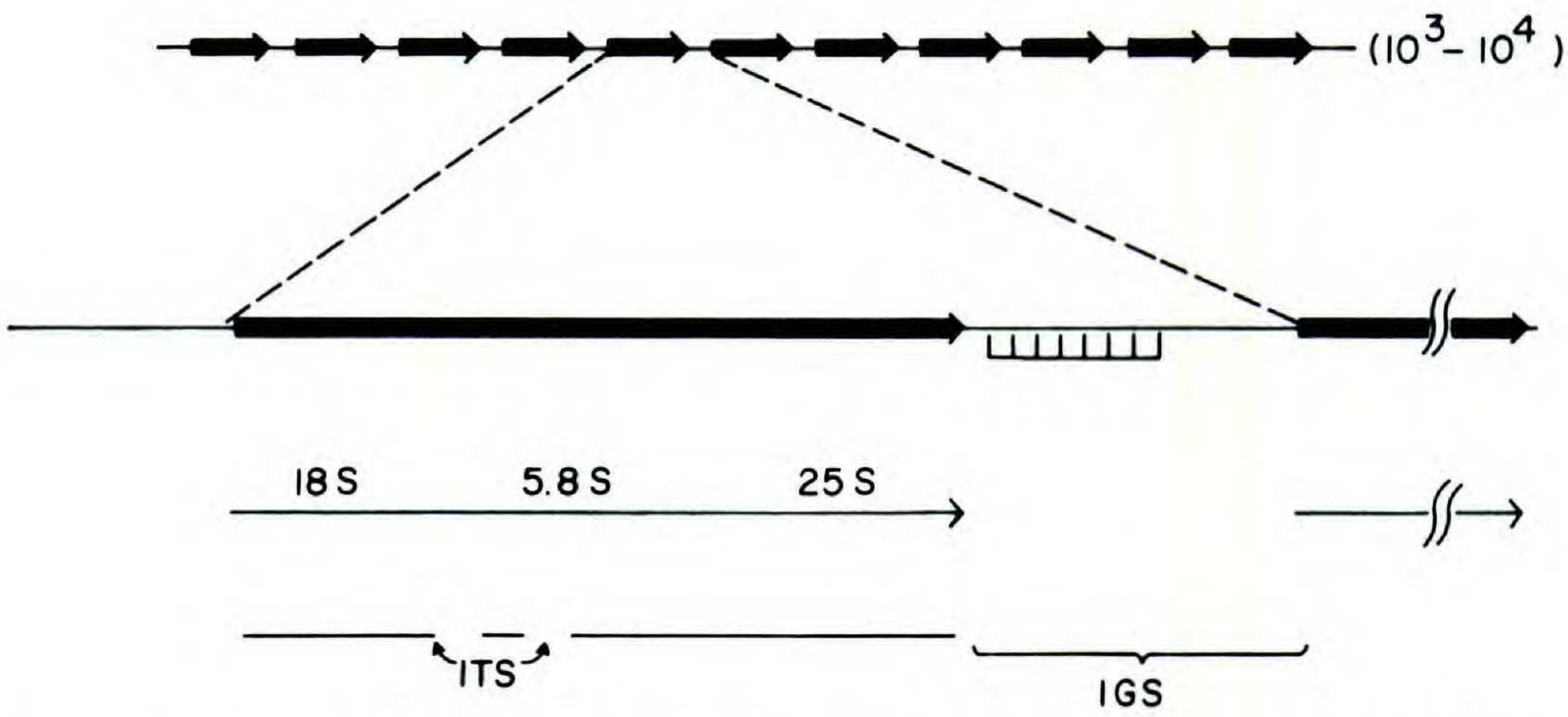

FIGURE 1. *Schematic representation of rDNA repeat structure. 18S, 5.8S, and 25S refer to ribosomal RNAs. ITS and IGS refer to internal transcribed spacer and intergenic spacer, respectively.*

anism responsible for variations in copy number (Szostak & Wu, 1980). Although there is some heterogeneity among copies of rDNA within individuals (see below), the rDNA repeat units of an individual plant are highly homogeneous. That is, while several types of rDNA repeat unit may be found in a single plant, many hundreds of repeat units are identical as assayed by Southern blot analysis. This homogeneity was first observed in comparisons within species contrasted with comparisons between species and is presumably the result of concerted evolution of rDNA repeat units, as explained by Arnheim et al. (1980). Similarly, in species where rDNA is found to reside at two or more genetic loci, repeat units are found to be quite homogeneous within each locus. Thus, loci can usually be distinguished by their repeat types, and homogeneity is greater within loci than between loci (Dvorak & Appels, 1982; Saghai-Maroof et al., 1984).

The physical structure of higher plant ribosomal DNA (Fig. 1) is similar to that in other higher eucaryotes (Long & Dawid, 1980, for review). The three ribosomal RNA coding regions lie in the order 5′, 18S, 5.8S, 25S, 3′, and are transcribed as a single large precursor, which is processed subsequently to the mature rRNA forms. Several hundred base pairs of DNA separate the 18S cistron from the 5.8S cistron and the 5.8S cistron from the 25S cistron. These two intercistronic regions are referred to as internal transcribed spacers (ITS). The region separating the transcription units of adjacent rDNA repeats is called the intergenic spacer (IGS, formerly NTS or nontranscribed spacer; Dover et al., 1982) and in most plants ranges in length from one to eight kilobase pairs (kb). A tandemly repeating sequence comprises part of the IGS region. This sequence varies interspecifically in length, ranging generally from 100 to 200 bp, while within species its length varies only slightly. The length of this subrepeat has been shown to be 130 bp in wheat, 180 bp in peas, 325 bp (comprised of two copies of a 155-bp sequence and one 14-bp sequence) in broad bean, 115 bp in both barley and oats, and 200 bp in maize (Appels & Dvorak, 1982; Jorgensen et al., 1982; Yakura et al., 1984; Saghai-Maroof et al., 1984; Cluster et al., in prep.; and McMullen et al., 1986, respectively). The number of these elements within a given rDNA repeat unit is variable, and thus the overall length of the IGS is variable, within and between populations. This variability in length of the IGS is discussed in detail below.

An individual plant's rDNA array is often heterogeneous with respect to the three basic modes of variation: length, nucleotide sequence, and base modification (e.g., Siegel & Kolacz, 1983; Appels & Dvorak, 1982; Waldron et al., 1983; Jorgensen et al., 1982, 1987). It should be noted that a fourth mode of rDNA variation occurs, namely variation in the copy number of rDNA per haploid genome; because it is a quantitative character it is rarely measured. rDNA copy number is unlikely to be informative taxonomically because it is extremely variable within species, although in genetic analyses it may have some utility. In Figure 2 the three principal modes of variation are illustrated for a single individual of the garden pea (Jorgensen et al., 1987). This individual carries a minimum of three types of rDNA repeat units, and each of the three is distinguished by the three modes. First, each type of repeat has a different overall length due to variation in the number of 185-bp subrepeats in the IGS. Second, nucleotide substitutions are evident in several regions of the rDNA repeat unit. Type "L" repeats (see Fig. 2) carry EcoRI sites in two of their nine subrepeats that are not present in types "S" or "C," and in the nonsubre-

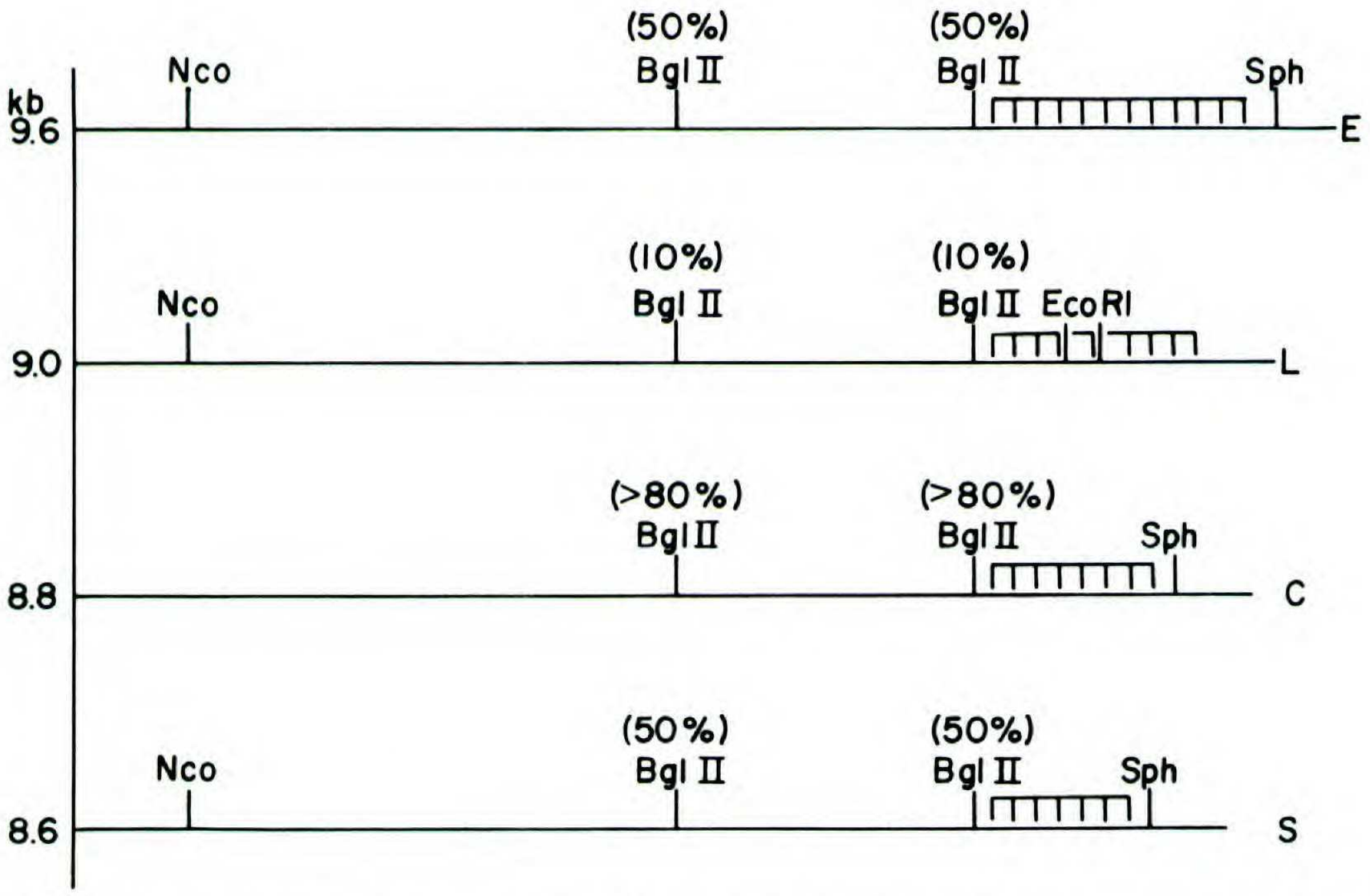

FIGURE 2. *Maps of four pea rDNA repeat units. L, C, and S refer to distinct repeat types found in a single pea plant. E refers to another repeat type found in* Pisum elatius. *Nco and Sph refer to NcoI and SphI.*

peat region of the IGS types "C" and "S" carry an SphI site not found in type "L." Within the rRNA coding sequence, an NcoI site in the 18S gene is present in both type "L"and "S," but absent in the cloned type "C" repeat. Third, the BglII cleavage sites of types "L"and "S" differ in their degree of apparent base modification such that only 10% of the BglII sites of "L" repeats are cleaved, while 50% of the BglII sites of "S" repeats are cleaved.

Temporal Analysis of Plant rDNA Variation

For a given mode of sequence evolution, the taxonomic level at which any segment of DNA is useful for making phylogenetic determinations is determined by the tempo at which that segment of DNA varies. Tempo can be estimated by analyzing DNA variation in species from several levels within an accepted taxonomic hierarchy. We have chosen nine legume genera for study: three (*Vicia*, *Pisum*, and *Lathyrus*) are from the tribe Vicieae and the rest (*Medicago*, *Trifolium*, *Lupinus*, *Wisteria*, *Cytisus*, and *Phaseolus*) are each from a different tribe. Seven genera are represented by a single species. *Vicia* is represented by five species, *Pisum* by four. For *V. sativa* and *P. sativum*, four and twenty, respectively, distinct isolates were examined. Outside the legume family we have compared the rDNA of wheat (*Triticum aestivum*) and pumpkin (*Cucurbita pepo*). Postulated phylogenetic relationships between species in this hierarchy are depicted in Figure 3.

Using the cloned rDNA repeat unit from *Pisum sativum* and detailed physical maps of both the cloned repeat and the pea nuclear genomic repeats described above, we performed nitrocellulose blot analysis (Southern, 1975) of rDNA sequences in the genomic DNA of each species. "Southern analysis" requires the use of a specific probe homologous to DNA sequences being analyzed. Different regions of the rDNA repeat unit were analyzed independently by use of seven different purified restriction fragments as probes (Fig. 4). The ITS region was analyzed by DNA sequencing because it was too small to analyze effectively by Southern analysis.

Modes and Tempos of Variation in Different rDNA Regions

A. Base Modifications

We have characterized in some detail two types of base modification in plant rDNA. They are distinguished both by sequence specificity and by degree of variability among taxa, as will be explained here. Most common is an evolutionarily conservative type of base modification typified by the BamHI site in the 25S gene. This site is modified in about one-half of the rDNA repeats of all the legume species in our survey, as well as in several other species (Gerlach & Bedbrook, 1979; Goldsborough et al., 1981; Jorgensen et al., 1982; Siegel & Kolacz, 1983). Siegel & Kolacz (1983) have postulated that this methylation is due to a CCG sequence of which the BamHI site (GGATCC) is a part. (That only one-half the sites are cleaved by

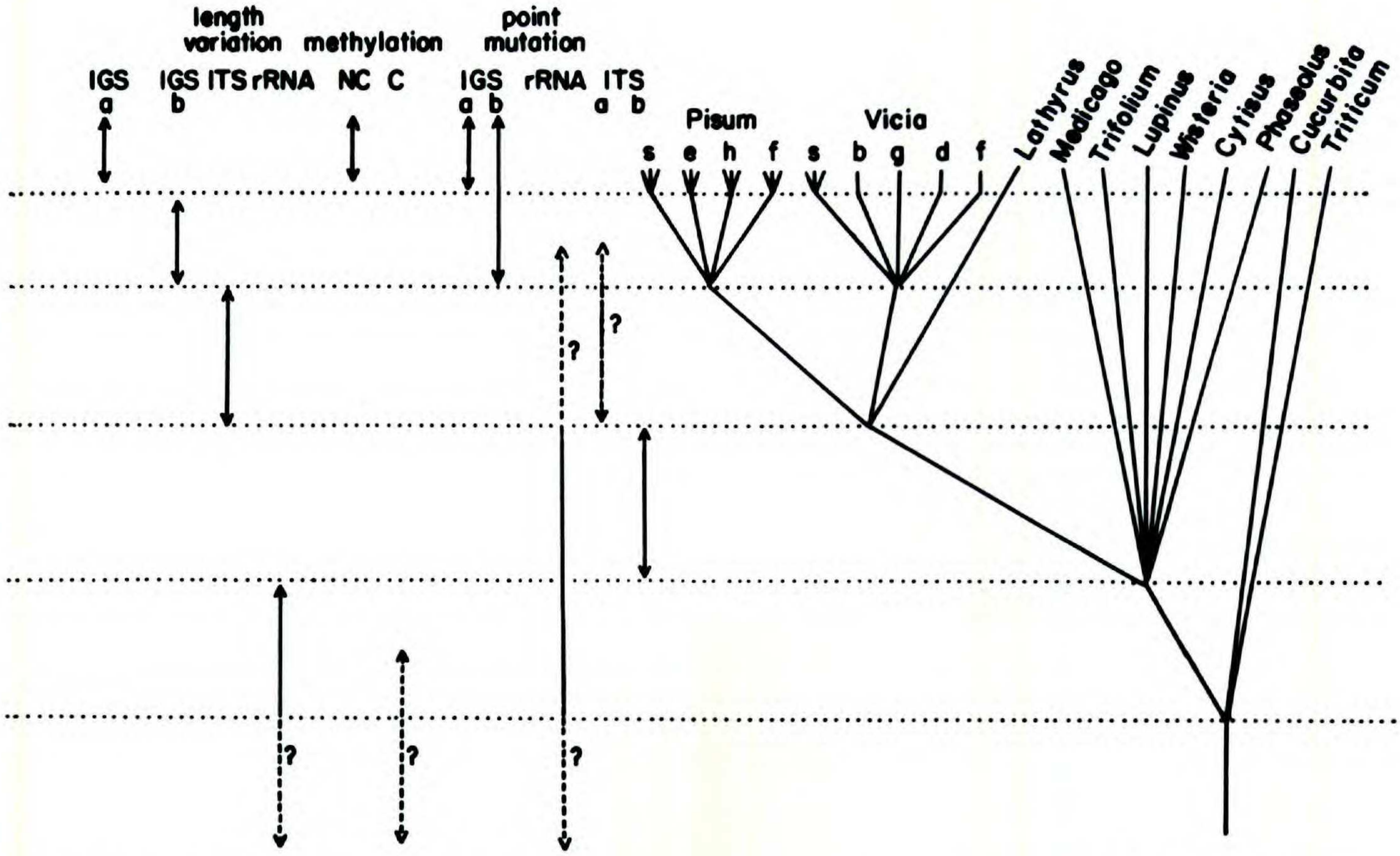

FIGURE 3. *Schematic drawing of evolutionary relationships among species analyzed and the taxonomic levels at which the eleven characters classes are proposed to be useful in reconstructing phylogeny. The taxonomic levels indicated are intraspecific [*Pisum sativum *(s)*, P. elatius *(e)*, P. humile *(h)*, P. fulvum *(f)*, *and* Vicia sativa *(s)]*, *intrageneric (*Vicia *and* Pisum*)*, *intratribal (Vicieae)*, *intrafamilial (Fabaceae)*, *and within angiosperms. The cladogram is not meant to represent relationships but is presented only to illustrate the taxonomic levels being compared. IGS refers to intergenic spacer; ITS refers to internal transcribed spacer; NC to nonconservative; C to conservative.*

BamHI appears to be due to the fact that methylation occurs at either but not both C residues.) Methylation of CG and CXG sequences has been observed in all plants investigated to date. It is presumed that many of the other enzyme cleavage sites that contain CG or CXG sequences could be subject to evolutionarily conservative modification. However, the extreme conservatism of the 25S BamHI site modification is a specific example of the general phenomenon of plant cytosine modification, and it is unwise to generalize from this. In fact, within or near various structural genes a substantial number (and perhaps a large fraction) of CG and CXG sequences are unmethylated in a variety of plants, and the possibility of variation in plant CG, CXG modifications certainly exists.

In contrast to the BamHI modification, modification of the BglII sites in pea rDNA is apparently variable by degree among individual pea plants, just as it is variable among rDNA repeats of the same plant (Jorgensen et al., 1982, 1987). BglII sites do not contain CG or CXG sequences but could be part of CXG sequences. It is not clear whether the variable modification of Bg1II sites in rDNA is due to variation in (a) modification by the CG, CXG system, (b) modification by another system, or (c) sequences adjacent to the site. Adenine modification prevents cleavage by certain restriction enzymes, but these have not been analyzed for variability among plants.

B. Single Base Pair Substitutions

1. Coding Regions. The coding regions for mature rRNAs were compared in two ways: by comparing restriction maps of cloned repeats from pea, wheat, and pumpkin (Jorgensen et al., 1987) and by comparing Southern blots of legume species rDNA using probes A, B, C, and D (Fig. 4). The 18S genes of pea and wheat were found to differ at three of ten six-bp cleavage sites (at least three of 60 bp), while the genes of wheat and pumpkin differ at five of nine, and the genes of pea and pumpkin at two of eight. The 5′ end of the 25S gene shows no site conservation in comparisons of these three species, which is consistent with the fact that this is one of the last-conserved regions in comparisons among frog, yeast, and slime mold rDNA (Gerbi et al., 1982). The rest of the 25S gene shows substantial similarity among species:

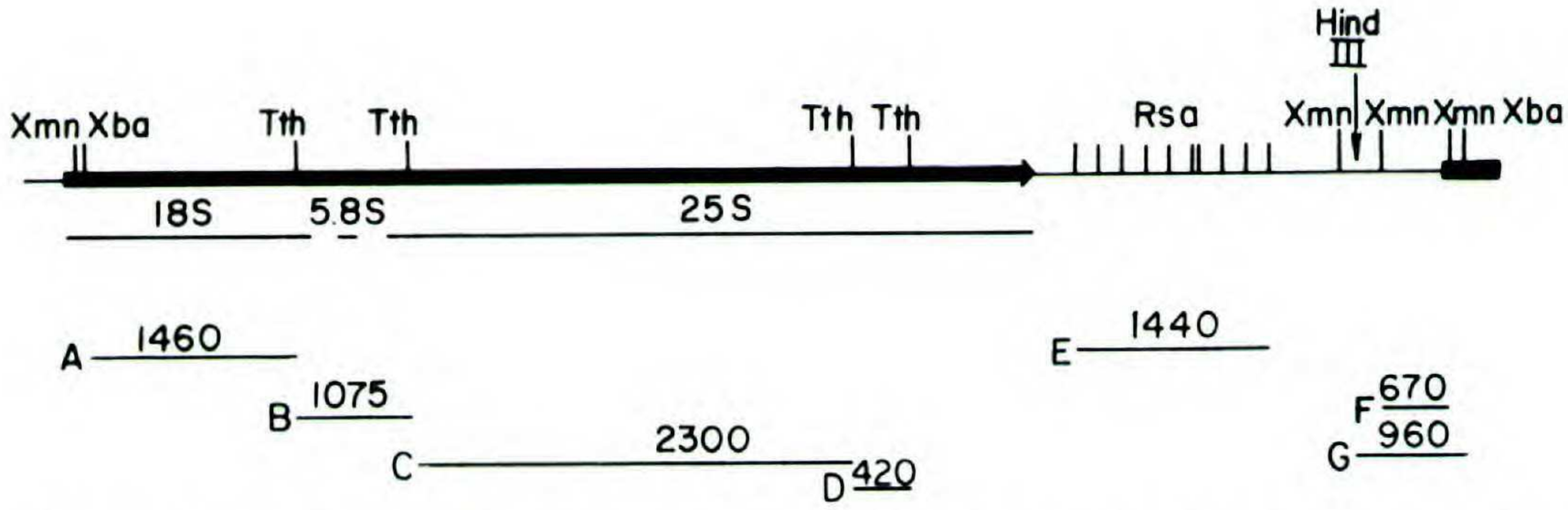

FIGURE 4. *Map indicating restriction fragments (A–G) used to study rDNA variation. Numbers indicate length of fragments in base pairs. Xmn, Xba, Rsa, and Tth refer to XmnI, XbaI, RsaI, and Tth111I.*

pea and pumpkin differ at three of ten sites, pea and wheat at eight of twelve, and wheat and pumpkin at five of ten. Although it is possible to align and compare the restriction maps of the pea, wheat, and pumpkin coding regions, a statistically significant number of mutations have not been analyzed to permit phylogenetic conclusions.

Southern analysis of different legume species reveals very little sequence divergence in the rRNA coding sequences, even after use of all available enzymes insensitive to base modification (Jorgensen et al., 1982). Figure 5 summarizes this analysis with a comparison of pea, vetch, and bean, illustrating the only two cleavage site mutations that could be detected in this survey of 19 cleavage sites (for a survey of 114 bp). The degree of sequence divergence in this rDNA region is at least several-fold less in the legume family than among peas, pumpkins, and wheat.

It is important to recognize the limitations inherent in the use of restriction enzyme analysis of plant rDNA for phylogenetic investigations. First, the choice of restriction enzymes for nuclear DNA analysis in plant genomes is more limited than for animal genomes or for plant organellar genomes due to the fact that plant nuclear DNA is methylated at most CG dinucleotides and CXG trinucleotides (Gruenbaum et al., 1981), and because many restriction enzymes that cleave sequences containing CG, CXG sequences do not cleave if these sequences are methylated. Thus, analysis of genomic rDNA by Southern blot is quite limited relative to analysis of rDNA clones using restriction enzymes with respect to the number of variants that can be detected. Cloned sequences are not necessarily a good alternative because the use of single clones from an array of thousands entails the risk of not being representative. Second, because rRNA coding sequences are only 5.5 kb long, in contrast to chloroplast DNA which amounts to about 150 kb (see Palmer et al., this volume), relatively few cleavage sites are available and relatively few variants can be detected. The obvious solution to this problem is to utilize rRNA sequencing. By doing so, Zimmer (this symposium, not published here) has shown that the conservative nature of the coding region is extremely useful in phylogenetic comparisons between distantly related genera and closely related families.

2. ITS Region. The ITS region, because it is small, also cannot be analyzed well with restriction enzymes. The DNA sequence of the 5.8S gene and its surrounding ITS sequences has been determined in pea and lupine (Jorgensen & Hess, unpubl.; Rafalski et al., 1983). A schematic comparison of

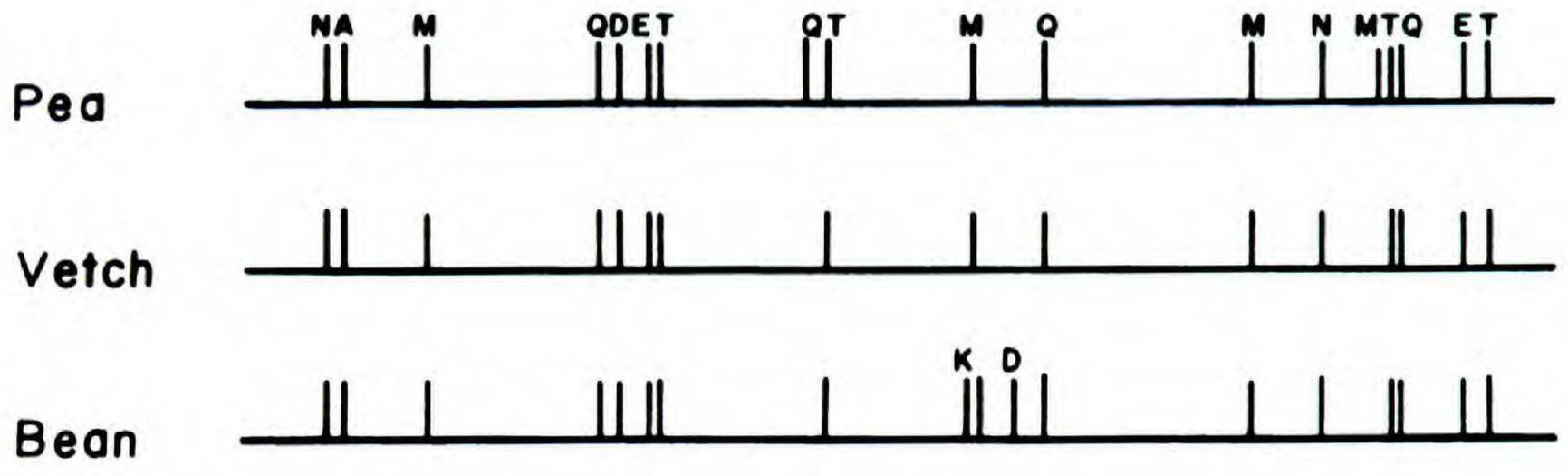

FIGURE 5. *Maps comparing coding regions of pea (*Pisum sativum*), vetch (*Vicia sativa*), and bean (*Phaseolus vulgaris*) rDNA. Symbols are coded as follows: N, XmnI; A, XbaI; M, BamHI; Q, SstI; D, BstEII; E, EcoRI; T, Tth111I; K, KpnI.*

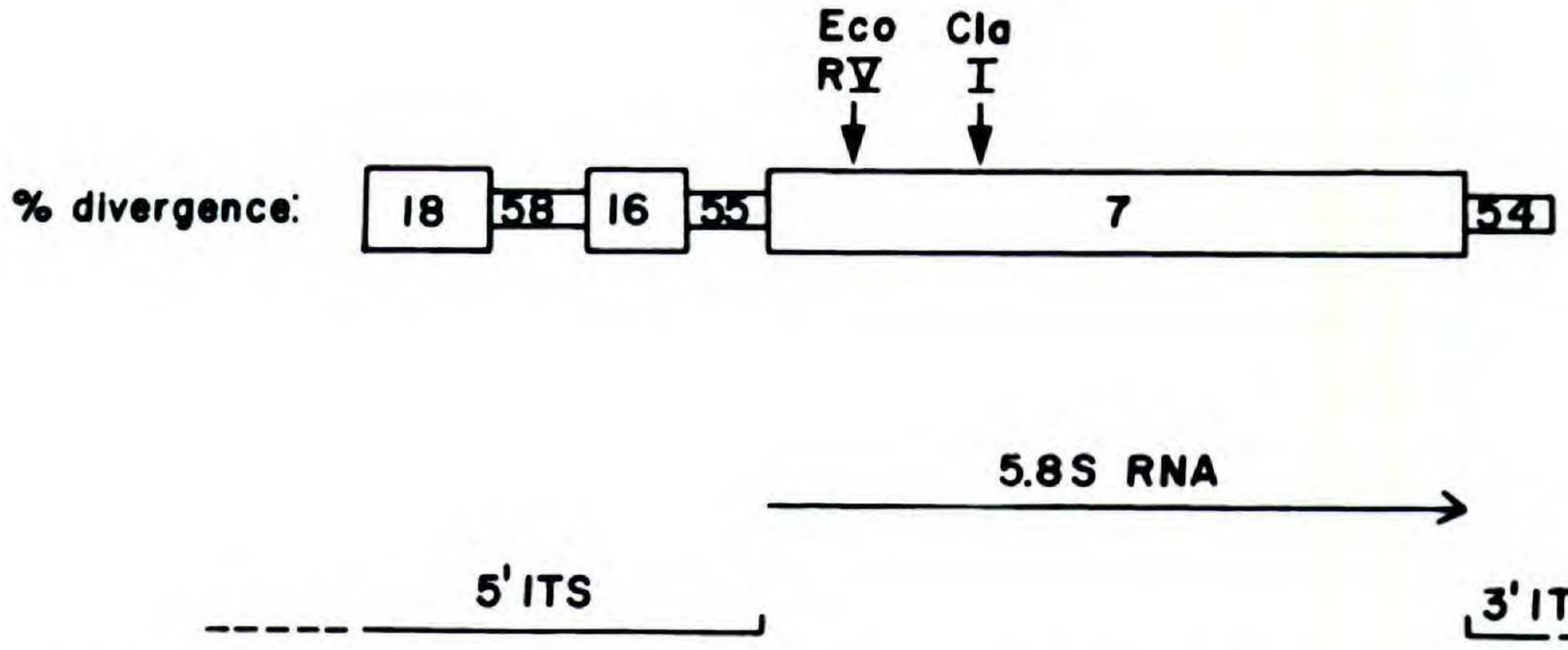

FIGURE 6. *Map of 5.8S rRNA gene and ITS regions. Numbers indicate % divergence in nucleotide sequence between* Pisum sativum *and* Lupinus luteus *rDNA clones.*

these sequences is shown in Figure 6. DNA in the ITS was observed to change at two distinct rates, both faster than that within the 5.8S gene. The 5.8S rRNA of *Vicia faba* has been sequenced (Tanaka et al., 1980) and differs from the pea 5.8S rRNA by 2%.

3. IGS Region. Restriction enzyme analysis of the nonsubrepeat segment of the IGS region reveals many cleavage site variants within legume genera (Jorgensen et al., 1982). These variants appear to have some utility in assessing phylogenetic relationships and in genetic studies of interfertile populations and species (Zimmer et al., in prep.). In the IGS subrepeats, DNA sequence and restriction pattern analyses indicate that subrepeat sequences are variable within and among individual genomes in a species (Appels & Dvorak, 1982; Jorgensen et al., 1987). This variation in single base pair substitutions in the IGS is of interest primarily within species. Because this variability is very difficult to detect with restriction analyses of total genomic DNA, it is of limited utility, except in studies of the evolution of the subrepeat itself.

C. Length Variation

1. Coding Regions. The ribosomal RNA transcription unit was surveyed for length variants by monitoring four restriction fragments (A–D) indicated in Figure 4. Fragment A, a 1,460-bp XbaI–TthI fragment, lies entirely within the 18S gene; fragments B (2,300-bp Tth fragment) and D (420-bp Tth fragment) lie entirely within the 25S gene; fragment C, the 1,075-bp TthI fragment, carries the entire ITS region and the 5.8S gene within it, as well as short portions of the 18S and 25S genes. Within the legumes, no length variations were observed (detection limit, 50 bp) in any of the three fragments that lie within the rRNA genes.

The 5.8S rRNA coding regions of pea, broad bean, and lupine have now been sequenced (Hess & Jorgensen, unpubl.; Tanaka et al., 1980; Rafalski et al., 1983). Comparison of these sequences shows that the 164-bp 5.8S gene differs in length between pea and broad bean by only one base pair and between pea and lupine by two adjacent base pairs. Length variation of this sort is very likely to be found also within the 18S and 25S genes by DNA sequence comparisons, but not by restriction fragment comparisons, as the differences are too small to detect by agarose gel electrophoresis.

2. ITS Region. Overall length variation in the ITS, as monitored by changes in the 1,075-bp TthI fragment, is much more prevalent than are length changes in the three coding sequence restriction fragments. The pea ITS 1,075-bp TthI fragment detects fragments varying in size from 1,000 to 1,200 base pairs among the nine legume genera surveyed. At least six of these size classes are distinct from all the others. Within *Pisum* and *Vicia* no length variation was observed. Thus, length variants of 50 bp or greater appear to be restricted mostly to the intergeneric level, at least in the tribe Vicieae. Whether small variants ever occur within these genera and whether observed variation results from the accumulation of many small variants or few large variants remains to be determined. For particular ITS length variants to be of practical use in studying relationships among genera, they will have to be large and rare, rather than small and common. Particular length variants in chloroplast DNA have been quite useful (in conjunction with point mutations) in developing chloroplast DNA phylogenies (Palmer et al., this volume).

3. IGS Region. By far the most variable region of the rDNA repeat unit is the subrepeat-containing region. Length variants of restriction fragments carrying this region almost always differ by a multiple of the length of the subrepeat. For

example, 14 distinct fragment lengths were observed in a sample of 12 pea lines, each length differing from the others by a multiple of 180 bp. Similarly, detailed intraspecific surveys of IGS variability among hundreds of individuals have demonstrated variation produced by 15 increments of 115 bp in barley (Saghai-Maroof et al., 1984) and by 17 increments of 115 bp in wild oats (*Avena barbata*) (Cluster et al., in prep.), indicating that the subrepeat length is 115 bp in each of these species. This large number of variants results in a large number of IGS phenotypes observed within and among populations (see below). For example, 11 distinct phenotypes were observed among the 12 pea lines.

The next most length-variable region is the part of the IGS region without subrepeats. We have monitored this region in *Pisum* with a 960-bp restriction fragment (G) produced by combined cleavage with HindIII and XbaI and with two XmnI fragments (E and F) of 310 bp and 670 bp (Fig. 4). Among ten pea lines that showed extensive variation in length of IGS-a, we found only two IGS-b types. These probably resulted from a single substitution event, this resulting in several restriction site differences as well as a 100-bp length difference. This finding illustrates the need for caution in interpreting restriction patterns and consideration of the possibility that restriction site variants in the IGS-b may be the result of deletions, insertions, or substitutions rather than point mutations. Interpretation of the molecular basis of mutations in this region in the absence of direct DNA sequence information is prone to error due to the small length changes potentially involved.

D. Summary of Tempos and Modes

Based on the results described above, Figure 3 illustrates the taxonomic levels at which the 11 identified character classes may have some utility in evolutionary genetic studies. It should be noted that characters in the single base pair substitution (point mutation) mode will be best detected by sequencing of rRNA or cloned rDNA, not by restriction analysis, as is demonstrated by Zimmer (this symposium, not published here). Clearly the size of the rRNA coding region indicates that this region will provide the greatest number of characters and so will be the most informative. Further, we would expect this region to be useful at levels ranging from the intergeneric to the interfamilial. It would be interesting to test the utility of the nonrepeated IGS region for intrageneric comparisons, although length variations (i.e., additions and deletions) might preclude this possibility if they are found to occur so frequently as to obscure sequence similarities. Characters in the base modification and length modes will be best detected by restriction analyses of total genomic DNA. It appears that character classes in these two modes will be useful primarily in intraspecific genetic studies, as described in the next section.

Variation Within and Among Populations of a Single Species

The rDNA spacer length (sl) phenotype of individual wild oat plants usually is comprised of 4–10 variants, out of 17 variants known in the species, often in widely varying copy numbers. From among over 500 individuals sampled, at least 40 distinct phenotypes were distinguished by scoring the most abundant sl variants (Cluster et al., in prep.). Variation of rDNA sl phenotypes among populations of wild oats in California closely tracks previously established patterns of differentiation identified by allozymes, morphological characters, and quantitative characters. Furthermore, the degree of variability suggests that it may be possible to identify and differentiate populations on the basis of rDNA variability alone nearly as accurately as with the available set of variable allozyme loci. Similarly, in 75 samples of barley, 15 distinct sl phenotypes could be distinguished (Saghai-Maroof et al., 1984). Most of these were comprised of two or three sl variants. The level of intraspecific polymorphism in the IGS is, therefore, extremely high. In the case of wild oats, the ability of sl variants to distinguish in detail among and between populations is probably the result of two genetic properties of rDNA variants in this species: (1) that these variants lie at a minimum of four independently segregating loci and (2) that each locus contains hundreds or thousands of repeat units which can be of more than one sl variant type.

Another result of genetic analysis of rDNA is the observation of nonrandom distribution of sl variants in several species, which suggests that genetic exchange occurs less frequently between than within nucleolus organizer regions. This situation occurs in barley (Saghai-Maroof et al., 1984), wheat (Dvorak & Appels, 1982), pea (Ellis et al., 1984; Polans et al., 1986), and mouse (Arnheim et al., 1982); however, random distribution has been reported in humans (Krystal et al., 1981). In wild oats the most abundant sl variants were present in nearly all isolates, including both parents of the single F2 analyzed, and so it is not possible to assess accurately whether these variants are

distributed randomly or nonramdomly among loci, even though five less-abundant variants are nonrandomly distributed. In species where a degree of nonrandomness is observed, rDNA appears to be a new and useful genetic marker (Saghai-Maroof et al., 1984). However, the multilocus nature of rDNA spacer length variation may place a severe limitation on its use in population genetics because of the difficulty in determining each plant's genotype, except for those genomes possessing only a single major nucleolus organizing region (e.g., tomato and corn). Still, the great amount of phenotypic diversity will clearly be useful.

Also noteworthy is the observation that the composite frequency distribution of rDNA sl variants in California wild oats shows a nearly Poisson distribution of sl variants centered at sl variant 8. There at at least three ways to explain this distribution. First, it could be the result of classical forces in population divergence such as genetic drift and/or selection on loci at or correlated with rDNA. Second, it could be the result of stochastic molecular processes, perhaps involving DNA replication or repair. Third, it could be determined by the function of the subrepeat.

The first explanation is based on the fact that in populations that reproduce substantially by selfing, a correlational structure is imposed on all components of the genome, allowing selection at each locus to affect allelic frequencies at all other loci in the genome. Thus, sl variant frequencies at rDNA loci must reflect the selective effects of many other loci and could be determined by these forces (Saghai-Maroof et al., 1984).

The second explanation considers whether the observed distribution could be simply a consequence of the molecular mechanisms that create new spacer length variants. New variants can appear in evolutionary time by mechanisms such as unequal crossing-over, resulting in repeated cycles of amplification and contraction of arrays of both repeats and subrepeats of rDNA. Accordingly, one can hypothesize that the number of subrepeats in the intergenic spacer region would be determined by a feature(s) of these mutational mechanisms whereby very long subrepeat arrays are more likely to be shortened than lengthened and short arrays are more likely to be lengthened than shortened, resulting in a balance in wild oats, for instance, at slv-8. Since various cellular processes might affect this mechanistic optimum, it could be possible for species to differ in their optimum number of subrepeats, were such a mechanism the only one operating on the distribution.

The third explanation considers what function the IGS might have, based on recent observations on the transcriptional and structural nature of the subrepeat elements (Reeder, 1984; Flavell et al., 1987). Briefly, it is hypothesized that the IGS subrepeats function analogously to enhancer sequences, increasing the transcription of the repeat unit(s) to which they are adjacent. It has been observed that rDNA repeat units with more subrepeats are transcribed with strong preference over units with fewer subrepeats, probably due to an interaction between subrepeats and some positive transcription factor. Furthermore, loci with repeat units having more subrepeats show nucleolar dominance over loci having fewer repeats. These observations provide a simple explanation of how natural selection for longer sl variants might occur. Of course, subrepeats apparently do not increase to many tens or hundreds of copies. Therefore, we must also hypothesize that too large a number of subrepeats can be deleterious. Perhaps multiple subrepeats would sequester a transcription factor not only from rRNA promoters in unlinked loci but also from promoters in adjacent repeat units (see Reeder, 1984, for detailed discussion of this model). This situation would likely lead to a reduction in efficiency of rRNA transcription and thereby be deleterious to the individual. Thus, it is possible that natural selection can *directly* mold the rDNA sl variant pattern and influence the frequency distribution seen in wild oats.

Summary

Eleven classes of useful characters have been identified for plant nuclear ribosomal RNA genes. These classes and their approximate rates of evolution are as follows:

(1) The length of the plant ribosomal DNA repeat unit is highly variable within most species and this variability has great utility in studies of populations. A 100–200-bp sequence that is repeated many times in tandem in the IGS region of rDNA forms the molecular basis for this variation in that the number of tandem copies of this sequence differs among individuals as well as among rDNA repeats within an individual. Studies of these variants appear to be helpful in elucidating the molecular mechanisms of rDNA evolution.

(2) The nonrepeating portion of the IGS region is less variable than is the subrepeat region, but is variable in length within the genera *Pisum* and *Vicia*; this variability might have utility in assessing specific relationships within such genera.

(3 and 4) Ribosomal RNA coding sequences are invariant in length in restriction pattern analyses,

but length variation is observed in the internal transcribed spacer (ITS) region. The frequency at which length variations in the ITS are detected is lower than that for length variation in either segment of the IGS. It is not likely that length variation in this region will be useful phylogenetically.

(5 and 6) Modification of Bg<II cleavage sites in pea rDNA varies quantitatively among pure lines and between wild populations. Such variants might be useful in genetic studies, but only to a limited degree since they are not as frequently observed as IGS length variants. A second class of modification is extremely conservative, as is illustrated by one of the rDNAs BamHI cleavage sites, which is modified in all legume and cereal species so far examined.

(7–11) Nucleotide substitutions accumulate in plant rDNA at several distinct tempos in distinct segments of DNA, resulting in a minimum of five character classes. In the IGS region they are found in (a) the subrepeat segment, differing among rDNA repeats of single individuals, and (b) in the nonrepeating segment within and among species of the same genus. Comparison of ITS DNA sequences of pea and lupine reveals four blocks of sequence, each of which has accumulated base substitutions at one of two levels, both greater than within the 5.8S rRNA coding sequences as determined by comparing sequences of pea, broad bean, lupine, and wheat and restriction enzyme cleavage sites of pea, broad bean, runner bean, pumpkin, and wheat.

Literature Cited

Appels, R. & J. Dvorak. 1982. The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. Theor. Appl. Genet. 63: 337–348.

Arnheim, N., D. Treco, B. Taylor & E. Eicher. 1982. Distribution of the ribosomal gene length variants among mouse chromosomes. Proc. Natl. Acad. U.S.A. 79: 4677–4680.

———, M. Krystal, R. Schmickel, G. Wilson, O. Ryder & E. Zimmer. 1980. Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. Proc. Natl. Acad. U.S.A. 77: 7323–7327.

Cullis, C. & D. R. Davies. 1975. Ribosomal DNA amounts in *Pisum sativum*. Genetics 81: 485–492.

Dover, G. A., S. D. M. Brown, E. S. Coen, J. Dallas, T. Strachan & M. Trick. 1982. The dynamics of genome evolution and species differentiation. Pp. 343–372 *in* G. A. Dover & R. B. Flavell (editors), Genome Evolution. Academic Press, London.

Dvorak, J. & R. Appels. 1982. Chromosome and nucleotide sequence differentiation in genomes of polyploid *Triticum* species. Theor. Appl. Genet. 63: 349–360.

Ellis, T. H. N., D. R. Davies, J. A. Castleton & I. D. Bedford. 1984. The organization and genetics of rDNA length variants in peas. Chromosoma 91: 74–81.

Flavell, R. B., M. O'Dell, P. Sharp, E. Nevo & A. Beiles. 1987. Variation in the intergenic spacer of ribosomal DNA of wild wheat, *Triticum dicoccoides*, in Israel. Molec. Biol. Evol. 3: 547–558.

Gerbi, S. A., R. L. Gourse & C. G. Clark. 1982. Conserved regions within ribosomal DNA: locations and some possible functions. Pp. 351–386 *in* The Cell Nucleus, Volume X. Academic Press, London.

Gerlach, W. L. & J. Bedbrook. 1979. Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res. 7: 1869–1885.

Goldsborough, P. B., T. H. N. Ellis & C. A. Cullis. 1981. Organization of the genes for ribosomal RNA in flax. Nucleic Acids Res. 9: 5895–5905.

Gruenbaum, Y., T. Naveh-Many, H. Cedar & A. Razin. 1981. Sequence specificity of methylation in higher plant DNA. Nature 292: 860–862.

Ingle, I., I. N. Timmis & J. Sinclair. 1975. The relationship between satellite desoxyribonucleic acid, ribosomal ribonucleic acid gene redundancy and genome size in plants. Pl. Physiol. 55: 496–501.

Jorgensen, R. A., R. E. Cuellar & W. F. Thompson. 1982. Modes and tempos in the evolution of nuclear-encoded ribosomal RNA genes in legumes. Carnegie Inst. Washington Year Book 81: 98–101.

———, ———, ——— & T. A. Kavanaugh. 1987. Structure and variation in ribosomal RNA genes of pea. Pl. Molec. Biol. 8: 3–12.

Krystal, M., P. D'Eustachio, F. H. Ruddle & N. Arnheim. 1981. Human nucleolus organizers on nonhomologous chromosomes can share the same ribosomal gene variants. Proc. Natl. Acad. U.S.A. 78: 5744–5748.

Long, E. O. & I. B. Dawid. 1980. Repeated genes in eucaryotes. Annual Rev. Biochem. 49: 727–764.

McMullen, M. D., B. Hunter, R. L. Phillips & I. Rubinstein. 1986. The structure of the maize ribosomal DNA spacer region. Nucleic Acids Res. 14: 4953–4968.

Polans, N. O., N. F. Weeden & W. F. Thompson. 1986. Distribution, inheritance, and linkage relationships of ribosomal DNA spacer length variants in pea. Theor. Appl. Genet. 72: 289–295.

Rafalski, J. A., M. Wiewjiorowski & D. Soll. 1983. Organization of ribosomal DNA in yellow lupine (*Lupinus luteus*) and sequence of the 5.8S RNA gene. FEBS Lett. 152: 241–244.

Reeder, R. H. 1984. Enhancers and ribosomal gene spacers. Cell 38: 349–351.

Saghai-Maroof, M. A., K. M. Soliman, R. A. Jorgensen & R. W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. U.S.A. 81: 8014–8018.

Siegel, A. & K. Kolacz. 1983. Heterogeneity of pumpkin ribosomal DNA. Pl. Physiol. 72: 166–171.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Molec. Biol. 98: 503–517.

Szostak, J. W. & R. Wu. 1980. Unequal crossingover in the ribosomal DNA of *Saccharomyces cerevisiae*. Nature 284: 426–430.

Tanaka, Y., T. Dyer & G. G. Brownlee. 1980. An

improved direct RNA sequence method: its application to *Vicia faba* 5.8S ribosomal RNA. Nucleic Acids Res. 8: 1259–1271.

WALDRON, J., P. DUNSMUIR & J. BEDBROOK. 1983. Characterization of the rDNA repeat units in the Mitchell *Petunia* genome. Pl. Molec. Biol. 2: 57–65.

YAKURA, K., A. KATO & S. TANIFUJI. 1984. Length heterogeneity of the large spacer of *Vicia faba* rDNA is due to the differing numbers of a 326 bp repetitive sequence element. Molec. Gen. Genet. 193: 400–405.